Syst. Biol. 67(6):1010–1024, 2018

DOI:10.1093/sysbio/syy024
Advance Access publication March 19, 2018

Dating the Species Network: Allopolyploidy and Repetitive DNA Evolution in American

JAMIE MCCANN[1], TAE-SOO JANG[2], JIRI MACAS[3], GERALD M. SCHNEEWEISS[1], NICHOLAS J. MATZKE[4], PETR NOVAK[3], TOD F. STUESSY[1,5], JOSÉ L. VILLASEÑOR[6], AND HANNA WEISS-SCHNEEWEISS[1,*]

[1]*Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria;* [2]*Department of Biology, College of Bioscience and Biotechnology, Chungnam National University, Daejeon 34134, South Korea;* [3]*Biology Centre, Czech Academy of Sciences, Institute of Plant Molecular Biology, Branišovská 31, CZ-37005, České Budějovice, Czech Republic;* [4]*Division of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, ACT 2601, Australia;* [5]*Herbarium and Department of Evolution, Ecology and Organismal Biology, 1315 Kinnear Road, The Ohio State University, Columbus, Ohio 43212, USA; and* [6]*Department of Botany, UNAM, Tercer Circuito s/n, Ciudad Universitaria, Delegación Coyoacán, MX-04510 México, D.F., México*
**Correspondence to be sent to: Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria. Email: hanna.schneeweiss@univie.ac.at.*

Received 4 July 2017; reviews returned 17 February 2018; accepted 15 March 2018
Associate Editor: David Tank

Abstract.—Allopolyploidy has played an important role in the evolution of the flowering plants. Genome mergers are often accompanied by significant and rapid alterations of genome size and structure via chromosomal rearrangements and altered dynamics of tandem and dispersed repetitive DNA families. Recent developments in sequencing technologies and bioinformatic methods allow for a comprehensive investigation of the repetitive component of plant genomes. Interpretation of evolutionary dynamics following allopolyploidization requires both the knowledge of parentage and the age of origin of an allopolyploid. Whereas parentage is typically inferred from cytogenetic and phylogenetic data, age inference is hampered by the reticulate nature of the phylogenetic relationships. Treating subgenomes of allopolyploids as if they belonged to different species (i.e., no recombination among subgenomes) and applying cross-bracing (i.e., putting a constraint on the age difference of nodes pertaining to the same event), we can infer the age of allopolyploids within the framework of the multispecies coalescent within BEAST2. Together with a comprehensive characterization of the repetitive DNA fraction using the RepeatExplorer pipeline, we apply the dating approach in a group of closely related allopolyploids and their progenitor species in the plant genus *Melampodium* (Asteraceae). We dated the origin of both the allotetraploid, *Melampodium strigosum*, and its two allohexaploid derivatives, *Melampodium pringlei* and *Melampodium sericeum*, which share both parentage and the direction of the cross, to the Pleistocene (<1.4 Ma). Thus, Pleistocene climatic fluctuations may have triggered formation of allopolyploids possibly in short intervals, contributing to difficulties in inferring the precise temporal order of allopolyploid species divergence of *M. sericeum* and *M. pringlei*. The relatively recent origin of the allopolyploids likely played a role in the near-absence of major changes in the repetitive fraction of the polyploids' genomes. The repetitive elements most affected by the postpolyploidization changes represented retrotransposons of the Ty1-*copia* lineage Maximus and, to a lesser extent, also Athila elements of Ty3-*gypsy* family.[Allopolyploidy; divergence time estimation; *Melampodium*; phylogenetics; repetitive DNA evolution; species network.]

Polyploidy plays an important role in eukaryotic genome evolution, especially in the plant kingdom (Madlung 2013; Weiss-Schneeweiss et al. 2013; Wendel 2015), and although the debate over its relevance for speciation continues (see Mayrose et al. 2011 and replies), it is clear that most plants stem from polyploid backgrounds (Comai 2005; Jiao et al. 2011). Allopolyploidy, in particular, combines hybridization and whole genome duplication (WGD) and is thought to be a mechanism contributing to diversification in plants (Grant 1981; Rieseberg and Willis 2007; Wood et al. 2009). This evolutionary process has also been shown to stimulate rapid and extensive genome reshuffling attributed to either hybridization, genome doubling or a combination of both (Koh et al. 2010; Barker et al. 2012).

Genome dynamics in allopolyploids typically reflect the processes of genetic and cytological diploidization (Wolfe 2001; Leitch and Bennett 2004; Ma and Gustafson 2005; Renny-Byfield et al. 2013; Hollister 2015). An important component of these dynamics is repetitive DNA, which is responsible for much of the genome size variation observed in the plant kingdom (Dodsworth et al. 2015). Repetitive DNA in plant genomes is composed largely of dispersed transposable elements (retrotransposons and DNA transposons; Bennetzen and Wang 2014) and tandem repeats, both noncoding (arrays of monomers of species- or genus-specific satellite DNAs; Macas et al. 2002; Garrido-Ramos 2015) and coding (ribosomal DNAs), arranged in distinct chromosomal loci (Kovařík et al. 2008). Major changes in the composition of repetitive DNA have been shown to occur soon after allopolyploidization in *Nicotiana* (Renny-Byfield et al. 2011, 2012, 2013), with a near-complete genome turnover occurring within a few million years only (Lim et al. 2007). Although changes in repetitive DNA landscapes on the genomic scale can now be comprehensively investigated due to technological (high-throughput sequencing) and analytical advances (dedicated bioinformatic pipelines, such as RepeatExplorer: Novák et al. 2010, 2013), comparative studies in allopolyploid species remain scarce (Renny-Byfield et al. 2011, 2012, 2013; Mandáková et al. 2013; Zozomová-Lihová et al. 2014).

Inferences of the dynamics and mechanisms of the evolution of polyploid genomes require understanding their origins, with respect to both their parentage

and age. The parental origin of an allopolyploid is typically inferred from a combination of morphological, cytogenetic, and molecular evidence. Hypotheses of parental origin can be tested and refined by genomic *in situ* hybridization (GISH; i.e., mapping of genomic DNAs of the putative parental taxa to allopolyploid chromosomes; Jang and Weiss-Schneeweiss 2015), additionally allowing for the assessment of the extent of interactions between the parental subgenomes in allopolyploids (Chester et al. 2012, 2015; Mandáková et al. 2013, 2014). Several phylogenetic methods for reconstructing species networks have been developed that can address, for instance, the assignment of allopolyploid homoeologues to their corresponding parental genomes and building the species networks from multilabeled trees (Than et al. 2008; Jones et al. 2013; Marcussen et al. 2012, 2015; Bertrand et al. 2015). A fully Bayesian approach incorporating assignment of all homoeologues and the multispecies coalescent to reconstruct allopolyploid species networks has recently been developed (AlloppNET and AlloppMUL models of Jones et al. 2013; Jones 2017), but it is currently available only for allotetraploids.

Despite these methodological advances in understanding allopolyploid origins, establishing an age for these origins remains problematic. Various aspects of the mechanisms of allopolyploid formation, including number of origins, the extinction of parental taxa (incomplete sampling) and the presence of multiple subgenomes in a single species complicate and bias a divergence time analysis (Doyle and Egan 2010). Bertrand et al. (2015) use the divergence times of parental and allopolyploid alleles in a simple Bayesian model to determine ages for the allopolyploidy events, but this may introduce bias as divergence times of genes do not necessarily correspond to those of the lineages (Kellogg 2016). This issue can be circumvented in the framework of the multispecies coalescent, where under the same assumptions made by most of the previous phylogenetic approaches applied to allopolyploids, that is, extant parental ancestors and disomic inheritance, the allopolyploid subgenomes may be treated as distinct "species" (Fig. 1). Dating a multilabeled tree obtained in this way, however, will result in independent (and likely different) age estimates for the splits of the allopolyploid subgenomes from their respective lower-ploid ancestors. However, we know that these "two" divergence events are really one allopolyploidization event, and so should have the same estimated age, even though the absolute age of this event is unknown (Fig. 1). A method to obtain a single age estimate, with credible intervals, for the splits between parental taxa and the corresponding allopolyploid subgenomes is "cross-bracing" the dating analysis. Cross-bracing puts a prior constraint on the age difference of nodes pertaining to the same event (Fig. 1). It was recently introduced in the context of gene duplications (Shih and Matzke 2013), and its utility in the context of allopolyploidy will be tested here.

An excellent group in which to perform dating of allopolyploid origins and to investigate repetitive DNA evolution following (successive) polyploidization is the cytologically diverse genus *Melampodium* (Asteraceae). Its approximately 40 species exhibit chromosome number variation due to dysploidy ($x = 9, 10, 11, 12, 14$) as well as polyploidy with 40% of the species being of polyploid origin (Stuessy et al. 2011; Weiss-Schneeweiss et al. 2012; McCann et al. 2016). The focal group of this study contains three diploid (*Melampodium americanum*, *Melampodium glabribracteatum*, and *Melampodium linearilobum*), one allotetraploid (*Melampodium strigosum*) and two allohexaploid species (*Melampodium pringlei* and *Melampodium sericeum*), all belonging to section *Melampodium*. The allotetraploid *M. strigosum* ($2n = 4x = 40$) originated from the hybridization of the diploids *M. americanum* ($2n = 2x = 20$) and *M. glabribracteatum* ($2n = 2x = 20$), and in turn has hybridized with the diploid *M. linearilobum* ($2n = 2x = 20$) to give rise to two allohexaploid species, *M. pringlei* ($2n = 6x = 60$) and *M. sericeum* ($2n = 6x = 60$; Weiss-Schneeweiss et al. 2012). These two allohexaploids not only share both parental species but also the direction of the cross (Weiss-Schneeweiss et al. 2012).

Here we use *Melampodium* to address how the age and parental origin of allopolyploids as well as ploidy level influence the dynamics of repetitive DNA evolution. Specifically, we test the hypotheses that (1) the extent of repetitive DNA composition divergence from parental taxa increases with time after allopolyploid origin and (2) that divergent evolution of ribosomal DNAs (rDNAs), both in terms of sequences and number of loci (Weiss-Schneeweiss et al. 2012), is paralleled by divergent evolution of the repetitive fraction of the genome. To this end, we (1) establish a dated phylogenetic

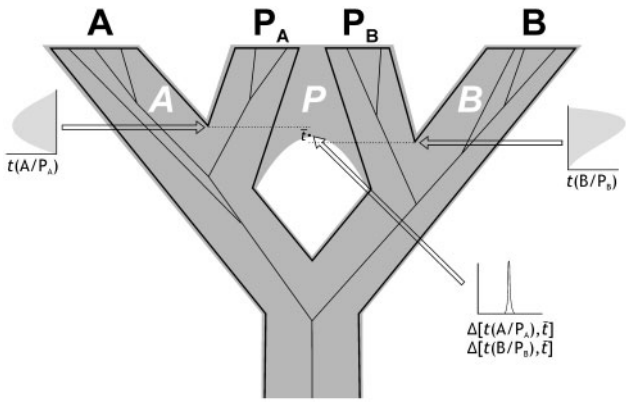

FIGURE 1. Application of cross-bracing to date the age of allopolyploids. Species ***A*** and ***B*** gave rise to the allotetraploid species ***P*** (gray network). Under the assumption that the parental subgenomes in the allotetraploid, $\mathbf{P_A}$ and $\mathbf{P_B}$, do not recombine, each is treated as distinct "species" (i.e., a gene pool/subgenome that has no history of intermixing with another gene pool/subgenome), resulting in a species tree with four leaves ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{P_A}$, $\mathbf{P_B}$). Split times between the lower-ploid species and the parental subgenomes [$t(A/P_A)$ and $t(B/P_B)$, indicated by dotted lines] have separate, explicit or implicit, priors (light grey distributions), but are forced by a narrow cross-bracing prior [white distribution; acting on the differences of the split age $t(A/P_A)$ and $t(B/P_B)$, respectively, from the mean split age $\bar{t}$] to be essentially contemporaneous, as they correspond to the same allopolyploidization event.

framework using cross-bracing in the program BEAST 2, (2) characterize the repetitive DNA in allopolyploid and parental genomes using the RepeatExplorer pipeline, and (3) investigate the dynamics of genome turnover in allopolyploids with respect to their parents and age of origin.

Materials and Methods

Dating Analyses

Molecular dating was done in two steps. First, the age of *Melampodium* sect. *Melampodium* was estimated using a larger data set from the tribe *Heliantheae* to which the genus *Melampodium* belongs. In the second step, the estimated age distribution for *Melampodium* sect. *Melampodium* was used as secondary calibration on the root node in a small data set including the focal allopolyploid species. All sequences were obtained from GenBank. Alignments (nexus files), BEAST XML files, and annotated MCC trees are deposited as Supplementary Material available on Dryad at http://dx.doi.org/10.5061/dryad.dg8q0 (see below).

Divergence time analysis for the genus Melampodium.—A collection of 159 sequences taken from lineages across the Heliantheae alliance, containing ITS1, 5.8S, and ITS2 of the 35S rDNA locus, were downloaded from GenBank (Supplementary Table S1 available on Dryad). Sequence alignment of this data set was performed with MUSCLE 3.8.31 (Edgar 2004) and further refined by eye. Partitioning schemes and best-fit substitution models were determined using the program IQ-TREE 1.3.4 (Nguyen et al. 2015) and chosen based on the Bayesian Information Criterion (BIC).

Molecular phylogenetic analysis and divergence time estimation were performed in BEAST 2.4.2 using a previous estimate on the age of the Heliantheae alliance (17–21 myr old; Kim et al. 2005, Torices 2010; see Supplementary Appendix S1 available on Dryad for details). A log-normal prior was used for the height of the root node. As a tree prior, a speciation model following a Yule process was used, with a diffuse inverse-gamma birth rate prior (alpha = 1, beta = 3). The monophyly of *Melampodium* (including *Acanthospermum* and *Lecocarpus*) and the monophyly of section *Melampodium* (excluding *Melampodium longipilum*) was enforced as suggested by previous work in Blöch et al. (2009).

Both a strict and an uncorrelated (log-normal) relaxed clock (Drummond et al. 2006) model were used with a gamma-distributed rate prior reflecting range estimates from Kay et al. (2006; see Supplementary Appendix S1 available on Dryad for details). Four separate runs of 25×10^6 generations (sampling every 10,000th generation) were performed for each model. The log files were merged without 10% burn-in for each run and examined for convergence [i.e., effective sample size (ESS) values of at least 200] using Tracer 1.8. Stepping stone sampling (Baele et al. 2012; 2013) was performed using an increasingly larger number of steps (always the same for both models) until the marginal likelihood estimates became stable. Each step was resumed until the ESS value of likelihood was >200. The marginal likelihood estimates for the competing clock models were used to choose the best-fit model according to the Bayes Factor.

Dating the age of allopolyploid origin in Melampodium sect. Melampodium.—One to two species were selected per phylogenetically and cytologically defined diploid genomic group (Weiss-Schneeweiss et al. 2012) that were considered most likely candidates involved in the allopolyploidy events (Weiss-Schneeweiss et al. 2012): *M. glabribracteatum* from diploid genomic group Glabribracteata; *M. linearilobum* from diploid genomic group Linearilloba, *M. americanum* and *Melampodium diffusum* from diploid genomic group Melampodium. The subgenomes of the allopolyploids were treated as separate "species." In this context, "species" does not refer to a taxonomic rank, but to a subgenome that has no history of intermixing with another subgenome, i.e., there is no recombination between subgenomes; within subgenomes, recombination between genes is permitted, while recombination within genes is not, thus following the same model assumptions used by Jones et al. (2013).

Sequences for five loci, the 5S rDNA nontranscribed spacer (NTS), the internal-transcribed spacers (ITS) 1 and 2 of 35S rRNA gene, the chloroplast gene *mat*K, and two paralogues of the nuclear gene *PgiC* (denoted I and II), for the selected diploid and allopolyploid species were downloaded from GenBank (accessions taken from Blöch et al. 2009; Weiss-Schneeweiss et al. 2012; Supplementary Table S2 available on Dryad), treated separately and aligned using MUSCLE 3.8.31 (Edgar 2004). Partitioning and substitution models were evaluated as outlined above for the Heliantheae alliance. The previous assignment of sequences from the allopolyploids to diploid genomic groups by Weiss-Schneeweiss et al. (2012) was re-evaluated using pairwise maximum likelihood (ML) estimates of genetic distance calculated in IQ-TREE 1.3.4 (Nguyen et al. 2015) for each locus. In case no sequence from a homoeologous subgenome was recovered (due to concerted evolution, gene loss, etc.), an empty sequence was used, as for estimation of all parameters in *BEAST (e.g., mean population size) all "species" (i.e., diploid species and subgenomes of allopolyploids) had to be represented at least once in each alignment and twice in at least one (Heled and Drummond 2010). For computational reasons (avoidance of many different stepping stone sampling runs), applicability of a strict molecular clock was tested for using likelihood ratio tests (Felsenstein 1981). To this end, maximum likelihood trees were calculated using IQ-TREE 1.3.4 (Nguyen et al. 2015), and likelihood scores were calculated on these trees with and without the molecular clock assumption in MEGA 7.0.1 (Kumar et al. 2016). As the simpler model of a homogeneous rate across the tree (i.e., a strict clock model) was significantly rejected in all cases ($P < 0.05$),

only relaxed clock models were used in the following BEAST analyses.

Relationships among "species" (i.e., diploid species and subgenomes of allopolyploids) and their divergence times were estimated using the multispecies coalescent as implemented in the *BEAST package of BEAST 2.4.2 (Heled and Drummond 2010; Ogilvie and Drummond 2017). In the species tree, the "species" corresponding to the allopolyploid subgenomes were forced to be monophyletic with their parental taxon. The split of the allopolyploid subgenomes from their lower-ploid ancestors was assumed to be (nearly) contemporaneous. This was achieved via the cross-bracing strategy of Shih and Matzke (2013). In contrast to cross-calibration, where the same prior distribution is applied to nodes of presumed same age (e.g., Marcussen et al. 2012), in cross-bracing a narrow, normally distributed prior (here, mean = 0 myr, standard deviation = 0.02 myr) is placed on the difference between the ages of the cross-braced nodes and the mean age of the cross-braced nodes (Fig. 1; for details including implementation in BEAST 2, which requires manual editing of the XML file, see Supplementary Appendix S2 available on Dryad). To account for the magnitude of the ages of the cross-braced nodes, the standard deviation of this prior is recommended to be chosen to be roughly in the order of 1% of the suspected age (Shih and Matzke 2013). This standard deviation allows the MCMC to move the ages of both nodes (a standard deviation of 0 would not), but ensures that the cross-braced node ages will be tightly correlated. This prior can be constructed to accommodate a single origin of an allopolyploid (permitting a difference of zero) as well as any desired length of time during which an allopolyploid may originate repeatedly, as is commonly observed in angiosperms (Soltis et al. 2010). Specifically, increasing the standard deviation of this cross-bracing prior allows the node height differences between the cross-braced nodes to behave more like non-cross-braced nodes (i.e., nodes without the cross-bracing prior; see Supplementary Appendix S2 available on Dryad).

For the two allohexaploid species, *M. pringlei* and *M. sericeum*, which have the allotetraploid *M. strigosum* as one of their parents (Weiss-Schneeweiss et al. 2012), all three scenarios of origin (a single origin vs. sequential origins of allohexaploids with *M. pringlei* splitting off first or second, respectively; see Supplementary Appendix S3 available on Dryad) were tested using Bayes Factors. Six separate runs of 10^8 generations for each scenario (sampling every 100,000th generation) were performed. After removal of 10% burn-in, the log files of each were combined and checked in Tracer for convergence (ESS > 200). Stepping stone sampling for determining the best scenario of origin for the allohexaploid species was performed using 112 steps and alpha = 0.3. Each step was run until the ESS values for the parameters were above 200. Absolute ages were obtained by putting a log-normal prior (mean = 5.5, standard deviation 0.2 in real space) on the root node, which was obtained from the age estimate of *Melampodium* sect. *Melampodium* in the first part of the divergence time analysis. To avoid overparameterization, a coalescent tree prior with constant population size was used for the gene trees, while the Yule prior was used for the species tree. Diffuse inverse-gamma prior distributions (alpha = 2, beta = 2) were used on both the population mean and the birth-rate parameters. The substitution model priors were left as the default settings for all loci. The species tree relaxed clock of Ogilvie and Drummond (2017) was used with rates drawn from a log-normal distribution, with log-normal and exponential hyperpriors on its mean (mean = 0.005, standard deviation = 0.35 in real space, thus accommodating rate variation reported for ITS sequences (Kay et al. 2006); see Supplementary Appendix S1 available on Dryad for details) and its standard deviation (mean = 0.33), respectively.

Plant Materials for NGS and Cytogenetic Analyses

Seeds, silica-dried leaves, and vouchers of all six species of *Melampodium* analyzed cytogenetically and for repetitive DNA composition were collected from natural populations in Mexico in August 2013 (collecting permit granted to J.L.V.). Herbarium specimens are deposited in the herbaria of the University of Vienna (WU), Ohio State University (OSU), and the National Autonomous University of Mexico (MEXU). Collection and accession numbers of each individual analyzed in this study are available in Supplementary Table S3 available on Dryad. Seed germination and plant cultivation were performed in the Botanical Garden of the University of Vienna (HBV).

DNA Isolation and Sequencing

Genomic DNAs (gDNAs) were isolated using a modified CTAB protocol (Doyle and Doyle 1987; Jang and Weiss-Schneeweiss 2015) and checked for quality and concentration using a Nanodrop spectrophotometer (PeqLab, Erlangen, Germany) and a fluorospectrophotometer and Quant-iT Picogreen dsDNA assay kit (PeqLab). DNA samples from two to three individuals per species, ideally from different populations, were pooled in equal proportions. Two independent libraries were prepared for each pooled sample (species), and these were sequenced separately. Fragmentation (600–800 nt in length) and library preparation were performed at the CSF-NGS sequencing facility (Vienna Biocenter, Austria). All samples were shotgun sequenced on a single lane of an Illumina HiSeq2500 machine (Illumina, San Diego, CA, USA) using the 150 nt paired-end technology. Genomic DNAs of putative parental taxa (Weiss-Schneeweiss et al. 2012) were also used for GISH.

Chromosome Spreads and GISH

Actively growing root meristems were harvested, pretreated with 0.002 M solution of 8-hydroxyquinoline for 2.5 h at room temperature and 2.5 h at 4°C, fixed in a 3:1 ethanol and acetic acid mixture, and stored

at −20°C until use (Weiss-Schneeweiss et al. 2012). Chromosome preparations were made after enzymatic digestion of fixed root meristems as described earlier (Weiss-Schneeweiss et al. 2012).

Genomic *in situ* hybridization was performed for the allotetraploid *M. strigosum* (M147, Hidalgo, Mexico) and for the allohexaploids *M. sericeum* (M63, Oaxaca, Mexico) and *M. pringlei* (M2089, Oaxaca, Mexico), using gDNAs of previously identified parental taxa (Weiss-Schneeweiss et al. 2012) as probes. Parental gDNAs of diploid *Melampodium linearilobum*, *M. glabribracteatum*, and *M. americanum*, as well as allotetraploid *M. strigosum* were sheared at 98°C for 5 min and labeled using either digoxigenin or the biotin nick translation kit (Roche, Vienna, Austria). The recently developed formamide-free hybridization and detection technique (Jang and Weiss-Schneeweiss 2015) was applied for GISH. Preparations were analyzed with an AxioImager M2 epifluorescent microscope (Carl Zeiss). Images were captured with a CCD camera and processed using AxioVision 4.8 (Carl Zeiss) with only those functions that apply to all pixels of the image equally.

Analysis of the Repetitive Fraction of the Genome using RepeatExplorer

Read pairs containing Illumina adapters, indeterminate bases (N) at any position, or failing to have a quality score $\geqslant$10 for at least 95% of the bases in either sequence were removed prior to clustering analysis using a combination of custom python and R scripts and the program BBMAP 34.65 (http://sourceforge.net/projects/bbmap/). Additional read filtering was performed to remove reads derived from plastid genomes and the *PhiX* spike-in DNA (Illumina). To that end, the sequence data sets were blasted against databases constructed from several Asteraceae plastid genomes (GenBank) and the *PhiX* genome. Read pairs were removed if both sequences produced blast hits with >90% sequence identity and over 90% of the read length. Filtering out mitochondrial reads was not possible at this stage due to the lack of a reference mitochondrial genome; as this affects all samples in the same way, no bias is expected, and clusters annotated as mitochondrial DNA were not counted as repetitive DNA in subsequent analyses. The remaining reads were trimmed to 140 nt by removing the first 10 nt of each sequence and assigned a unique three-letter species identifier for use in comparative clustering.

The reads were analyzed using the command line implementation of the RepeatExplorer pipeline (Novák et al. 2010, 2013). In brief, an all-to-all blast comparison was performed to generate a graph where the reads are connected by edges weighted by read similarity. The graph is then partitioned into highly interconnected sections which represent different repeat families (Novák et al. 2010). The identity of these families was determined by blasting the reads within a family to the default database of the plant transposable element domains included in the RepeatExplorer pipeline. The reads were analyzed using the default settings for sequence similarity and alignment length thresholds (Novák et al. 2013).

Clustering analyses were first performed individually for each species using the maximum number of reads possible for 100 GB of RAM (automatically estimated by RepeatExplorer and highly dependent on the species analyzed). Reads of two independently sequenced libraries of each taxon were analyzed together to detect biases due to library preparation. Clusters containing at least 0.01% of the total reads analyzed were manually annotated using both graph and dot-plot structure, and blast hits to the transposable element domain databases available by default in RepeatExplorer (Novák et al. 2013). Clusters containing reads that generated a cumulative number of blast hits to one or more protein-coding domains from the plant transposable element database >5% of the number of reads in the cluster were annotated with the lineage found in the database. Paired-end reads were used to annotate clusters unable to be annotated by other means when the ratio of number of pairs shared between clusters to the sum of the total number of missing mates from each cluster was >0.10 (Novák et al. 2013).

Proportions of solo-LTRs (long-terminal repeats) for each species were estimated as presented in Macas et al. (2015). This approach uses a blast database built from sequence tags at the junction of the 3′ end of the LTRs and the 5′ end of the untranslated region (UTR). The ratio of reads having blast hits to both regions to reads having hits only to the 3′ LTR region was used as an estimator of the ratio between solo-LTRs and full-length elements.

Following the individual species analyses, a number of reads from each species, scaled down to represent 0.1 × coverage of each genome (genome size data from Weiss-Schneeweiss et al. 2012), were sampled from only the forward mates of read pairs used in each of the individual analyses and combined into a single data set for comparative analysis using RepeatExplorer. This allowed for a more representative sample of the genome for a smaller number of reads. The resulting clusters in the comparative analysis, representing the same repeat types/families across the analyzed genomes, were annotated automatically using the cluster annotations from the individual species analyses.

To assess the similarity of the repetitive fraction of the genomes, a table was constructed with the number of reads derived from each species for each dispersed repeat-containing cluster obtained in the comparative analysis. Pairwise scatterplots were constructed for species on the same ploidy level, where the position of a single point in the scatterplot represents the number of reads in a cluster for the species on the x and y axes. Additionally, *in silico* allopolyploids, representing the expected number of reads given the additivity assumption, were constructed by summing the number of reads for the parental taxa in each row. These *in silico* allopolyploids were plotted against the observed number of reads for the actual allopolyploids in the

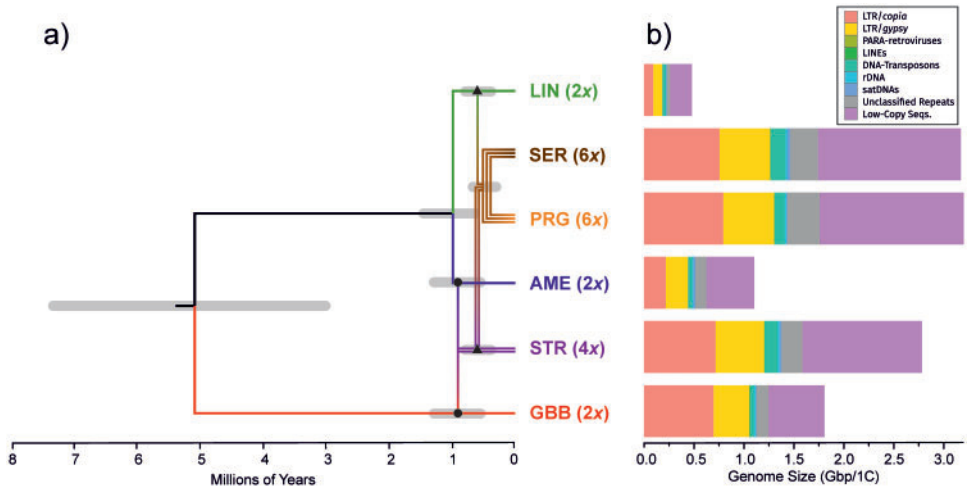

FIGURE 2. Phylogenetic relationships and repeat composition in analyzed *Melampodium* species. a) Dated phylogenetic network showing relationships between allopolyploids and parental taxa for the single allopolyploid origin scenario. Divergence times are shown at the nodes with 95% highest-posterior density intervals indicated by the gray bars; cross-braced nodes are marked by dots and triangles, respectively. b) Genomic proportions of repetitive elements for each species according to genome size and colored by repeat type (see legend). The species are abbreviated as follows: AME = *M. americanum*; GBB = *M. glabribracteatum*; LIN = *M. linearilobum*; PRG = *M. pringlei*; SER = *M. sericeum*; STR = *M. strigosum*. Ploidy level of each species is indicated as 2*x*, 4*x*, and 6*x* for diploids, tetraploids, and hexaploids, respectively.

same way as the pairwise scatterplots for the species on the same ploidy level. One-to-one lines were plotted in all scatterplots, where the slope of the line represents the ratio of the genome sizes between the species on the *x* and *y* axes. Deviation from this line represents differences in genomic proportion of a given cluster of reads for the species shown. In the allopolyploids, the slope of this was taken to be one, representing the additivity expectation of the allopolyploid subgenomes with respect to their parental taxa.

RESULTS

Dating the Age of Origin of Allopolyploids

For the Heliantheae data set, the alignment of the ITS1-5.8S-ITS2 region was 735 nt long (including gaps) with 496 variable sites. The best partitioning scheme according to the BIC was one in which the whole region was in a single partition with the TN +Γ substitution model with four discrete rate categories. Bayes Factors, estimated from the marginal likelihoods, indicated strong support for a relaxed clock model over the strict clock model (71.766). The mean tMRCA for *Melampodium* (including *Acanthospermum* and *Lecocarpus*) was 9.9 myr (95% HPD interval 7.5–12.7). The mean tMRCA age for *Melampodium* sect. *Melampodium*, comprising the focal group, was 5.1 myr (95% HPD interval 3.4–6.8 Ma; Supplementary Fig. S1 in Appendix S1 available on Dryad).

For the *Melampodium* sect. *Melampodium* data set, the amount of missing sequences ranged from 1.3 % (*mat*K) to 13.1 % (5S rDNA NTS spacer; Supplementary Table S4 available on Dryad). Failure to recover sequences from the homoeologous subgenome affected all allopolyploids, especially for plastid *mat*K and nuclear ITS sequences (Supplementary Table S4 available on Dryad). The strict clock hypothesis was rejected for ML trees obtained from all loci and corresponding sequence data sets (Supplementary Table S5 available on Dryad). Therefore, only the relaxed clock results are presented here. Bayes factors were inconclusive (maximally 0.8) with respect to the three alternative scenarios (maximum clade credibility trees are shown in Supplementary Figs. S5–S7 in Appendix S3 available on Dryad) for allopolyploid formation corresponding to a single origin for the allohexaploid or two possible scenarios for independent origins (Supplementary Table S6 available on Dryad).

Regardless of the scenario under consideration, the mean age of the ancestor to allotetraploid *M. strigosum* (i.e., the ages of the cross-braced nodes indicated with a black dot in Fig. 2a) was always around 0.9 million years ago (Ma; Fig. 2a, Supplementary Table S6 available on Dryad), ranging from 0.898 to 0.952 Ma across all scenarios. The 95% HPD estimates were 0.544–1.361 Ma, 0.579–1.414 Ma, and 0.561–1.308 Ma for the single origin (Supplementary Table S6 available on Dryad), *M. pringlei* first and *M. sericeum* first scenarios, respectively. Under the shared parental origin scenario, the allohexaploid ancestor of *M. pringlei* and *M. sericeum* (i.e., the ages of the cross-braced nodes indicated with a black triangle in Fig. 2a) had a mean age of 0.584 Ma (95% HPD interval 0.349–0.860 Ma; Fig. 2a). For the *M. pringlei* first and *M. sericeum* first scenarios, the mean ages of *M. pringlei* were 0.639 (95% HPD interval 0.354–0.947 Ma) and 0.506 Ma (95% HPD 0.294–0.750 Ma), while the mean ages of *M. sericeum* were 0.493 (95% HPD 0.257–0.749 Ma) and 0.571 (95% HPD 0.344–0.815 Ma), respectively (Supplementary Table S6 available on Dryad).

The differences in the ages of the cross-braced nodes representing the origin of the parental subgenomes were small in accordance with the cross-bracing prior

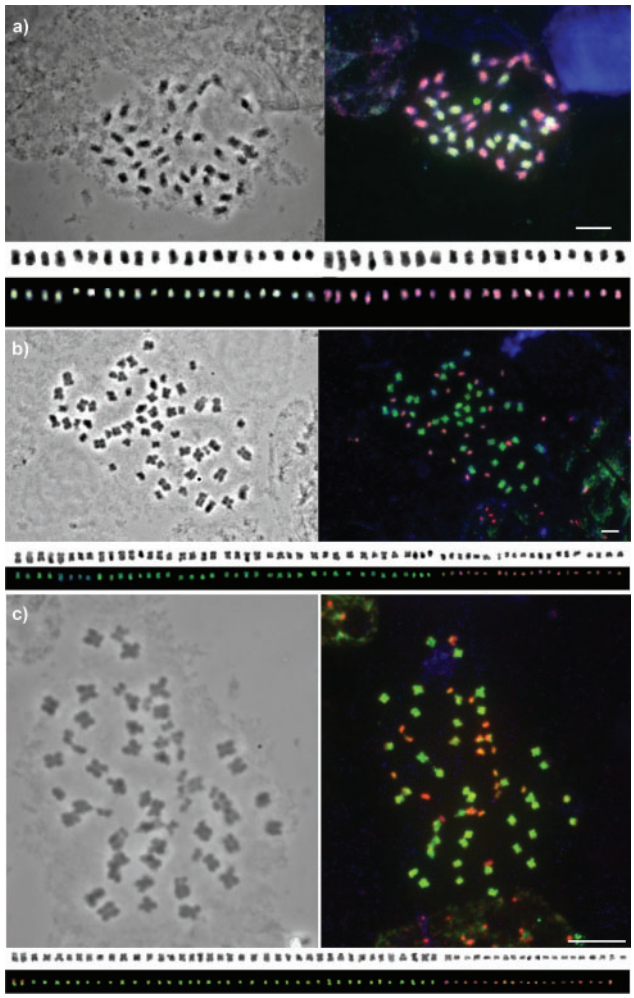

FIGURE 3. Phase contrast chromosomes (unstained), GISH and cut-out karyotype on mitotic chromosomes of the three allopolyploids. a) Allotetraploid *Melampodium strigosum* with labeled genomic DNA of diploid *Melampodium glabribracteatum* (green) and diploid *Melampodium americanum* (red). b, c) GISH of allohexaploid *Melampodium sericeum* (b) and of allohexaploid *Melampodium pringlei* (c) with labeled genomic DNA of diploid *M. linearilobum* (red) and allotetraploid *M. strigosum* (green). Scale bar, 5 μm.

imposed on these nodes. Individual BEAST runs needed upwards of half a billion generations to attain effective sample sizes >200 for each parameter value. Extending the run length to 2 billion generations (sampling every 100,000th generation) increased the ESS values, the lowest being those of the mean ages of the cross-braced nodes (ESS >975), but did not change the parameter estimates. Each of the runs of 600 million generations required slightly <2 days ($\sim$46 h) to complete on a computer with a 3.6 GHz processor, NVIDIA graphics card (GeForce GT 520) and the Beagle library installed (Ayres et al. 2012). To determine the effect of cross-bracing on the efficiency of MCMC exploration the probability of acceptance of node change proposals, affecting node height, was assessed. If proposed for a cross-braced node, the new node height may conflict with the cross-bracing prior, causing this node change proposal to be rejected. This will result in an overall reduced probability of acceptance of node change proposals. We expect that sampling efficiency will be decreased proportional to the number of cross-braced nodes in the tree. Indeed, the probability of node change proposals across the whole tree throughout the MCMC of the cross-braced runs was around 7.1% for all scenarios compared to 18.71% in the run without cross-bracing.

Parental Genome Identification in Allopolyploids

GISH with labeled gDNAs of putative diploid parental species, *M. americanum* and *M. glabribracteatum*, allowed unambiguous identification of the two parental chromosome sets in the allotetraploid *M. strigosum* ($2n=4x=40$) with 20 chromosomes each from *M. glabribracteatum* (green) and *M. americanum* (red; Fig. 3a). No intergenomic translocations were detected among the four tetraploid individuals analyzed (data not shown).

Genomic DNAs of the putative parental species, allotetraploid *M. strigosum* and diploid *M. linearilobum*,

TABLE 1. Species and DNA sequence information (from NGS data) for individual and comparative RepeatExplorer analyses

Species	ID	Ploidy level	Genome size[a]	Individual clustering		Comparative clustering
			Gbp/1C	No. reads	coverage	No. reads (0.1 ×)
Melampodium americanum	AME	2*x*	1.11	7621416	0.96 ×	897117
Melampodium glabribracteatum	GBB	2*x*	1.81	5555528	0.43 ×	1465723
Melampodium linearilobum	LIN	2*x*	0.48	7957510	2.32 ×	388729
Melampodium strigosum	STR	4*x*	2.79	5946104	0.30 ×	737049
Melampodium pringlei	PRG	6*x*	3.21	5787404	0.25 ×	721808
Melampodium sericeum	SER	6*x*	3.18	4568954	0.20 ×	729206

[a]From Weiss-Schneeweiss et al. (2012).

TABLE 2. Estimates of the genome proportion (%) of various repeat types identified in the analyzed diploid and allopolyploid genomes of *Melampodium*

Type	Repeat family	AME	GBB	LIN	STR	PRG	SER
Retrotransposons	—	39.96	58.92	37.95	43.49	40.74	39.97
copia	—	19.94	38.47	19.55	26.69	24.83	23.97
	Maximus	18.41	35.46	17.94	25.23	23.41	22.48
	Other	1.52	3.01	1.61	1.46	1.42	1.49
gypsy	—	20.03	20.46	18.40	16.80	15.91	16.00
	Athila	8.79	9.96	8.41	8.07	7.71	7.56
	Chromo	7.36	7.23	6.01	5.49	4.97	5.40
	Ogre/Tat	3.88	3.26	3.98	3.24	3.22	3.04
Other/nonLTR	—	0.69	0.81	0.73	0.52	0.59	0.50
	LINE	0.03	0.35	0.06	0.09	0.09	0.09
	SINE	0.14	0.04	0.14	0.11	0.12	0.13
	MITE	0.38	0.14	0.40	0.19	0.25	0.17
	PARA	0.14	0.27	0.13	0.14	0.13	0.10
DNA transposons	—	3.99	2.05	7.25	4.81	3.24	4.86
	CACTA	3.23	1.73	6.39	4.34	2.64	4.38
	Other	0.76	0.32	0.86	0.47	0.60	0.48
Tandem repeats	—	3.14	1.56	2.88	1.74	1.60	1.68
	rDNA	1.01	1.03	0.88	0.62	0.57	0.54
	satDNA	2.12	0.53	1.99	1.11	1.03	1.14
Unclassified	—	10.15	5.55	8.33	8.02	10.17	9.36
Total repeats	—	57.94	68.89	57.13	58.58	56.34	56.37
Low copy	—	42.06	31.11	42.87	41.42	43.66	43.63

AME = *M. americanum*; GBB = *M. glabribracteatum*; LIN = *M. linearilobum*; STR = *M. strigosum*; PRG = *M. pringlei*; SER = *M. sericeum*.

were mapped to chromosomes of the two allohexaploids, *M. pringlei* and *M. sericeum* (both $2n = 6x = 60$). In both hexaploid karyotypes, 20 chromosomes were clearly labeled with the genomic DNA of *M. linearilobum* (red) and 40 as *M. strigosum* (green; Fig. 3b,c). Two to four reciprocal translocations were detected in two of five individuals of *M. pringlei*, but no translocations were found in *M. sericeum* (Fig. 3b,c). To test the tri-parental origin of allohexaploids and determine the extent of genome homogenization within the *M. strigosum* parental subgenome in the two allohexaploids, a tri-color GISH experiment was performed (Supplementary Appendix S4 available on Dryad). The *M. americanum* and *M. glabribracteatum* subgenomes were labeled in the karyotypes of both allohexaploid species, *M. pringlei* and *M. sericeum*, as efficiently as in their allotetraploid parent *M. strigosum* and the 20 chromosomes of *M. linearilobum* were weakly labeled with the genomic DNA of *M. americanum* (Supplementary Fig. S8 in Appendix S4 available on Dryad).

Quantification of Repeats in Individual Genomes

Whole genome shotgun sequencing using the Illumina technology generated between 8 and 12 million 150 nt paired-end reads per sample (Table 1). Preprocessed data sets used for the final analyses were submitted to the NCBI Short Read Archive (SRA accession: SRP132795).

Clusters containing at least 0.01% of the total reads, corresponding to the medium- to high-copy number repeat families, were characterized in all species analyzed. Of these repeat families, 81–92% were successfully annotated. The largest proportions of unclassified clusters were found in the two allohexaploid species and in the diploid *M. americanum* (Table 2 and Fig. 2b). Repeats identified as retroelements were the most abundant, ranging from 37% to 59% of the total genomic composition in the diploid genomes and constituting ~40% of the allopolyploid genomes (Table 2). The majority of these repeats originated from the Maximus/SIRE lineage of the Ty1-*copia*

TABLE 3. Ratio of blast hits to the 3′-LTR (long terminal repeat)/5′-UTR (UnTranslated Region) junction (LU) to blast hits of the 3′-LTR/insertion site junction (LX) calculated as (LX − LU)/LU

Species	Athila	Angela	Ivana	Maximus	Tork	Athila	Chromovirus
AME	0.73	0.33	0.49	0.14	0.21	0.77	0.80
GBB	1.00	0.13	1.64	0.08	0.19	0.59	0.61
LIN	0.28	0.36	0.45	0.14	0.12	0.84	0.89
STR	0.75	0.33	1.07	0.11	0.36	0.67	0.67
PRG	0.58	0.55	0.85	0.07	0.33	0.75	0.94
SER	0.27	0.43	0.83	0.10	0.52	0.66	0.92

AME = *M. americanum*; GBB = *M. glabribracteatum*; LIN = *M. linearilobum*; STR = *M. strigosum*; PRG = *M. pringlei*; SER = *M. sericeum*.

retrotransposons, which comprised between 18% and 36% of the diploid genomes (*M. linearilobum* and *M. glabribracteatum*, respectively). Traces of the other main Ty1-*copia* lineages (Maximus, Ivana, Angela, Tork, Bianca, Tar, Ale-I) except Ale-II (Table 2) were also detected in all *Melampodium* species. All three major lineages of the Ty3-*gypsy* retrotransposons (in descending abundance Athila, Chromovirus, and Ogre/Tat) were found in *Melampodium*. Estimates of the ratios of solo-LTRs to full-length retroelements were $\leqslant 1$ for all repeat types across all species (Table 3).

DNA transposons were present in moderate amounts across all genomes, most of which were identified as *CACTA* type, with trace amounts of other lineages (Table 2). Other types of dispersed repeats detected included non-LTR retrotransposons (SINEs and LINEs) and para-retroviruses, all of which were found in trace amounts. Tandem repeats comprised relatively small proportions of the genome, but one tandem repeat, a microsatellite (ATTC) was abundant (>1% of the genome in *M. americanum*) in all species except *M. glabribracteatum* (Table 2).

Comparative Analysis of Repeat Dynamics in Allopolyploids and Their Parental Taxa

The identification of shared repeat families between three diploids, an allotetraploid and two allohexaploid species was performed using a comparative clustering approach. Approximately 9 million reads were analyzed in total, which amounted to about 0.1 × coverage for all species analyzed. Over 7 million reads were found in 378 clusters which contained at least 0.01% of all sequences analyzed.

The comparative analysis revealed considerable variation among diploids in dispersed repeat clusters (Fig. 4), particularly between *M. glabribracteatum* and the other two diploid species (*M. americanum* and *M. linearilobum*). The largest differences, based on the ratio of the genome sizes (red lines in Fig. 4), among these species were in clusters identified as Ty1-*copia* Maximus/SIRE repeats. Most of these repeats had higher copy numbers in at least one diploid species, but the vast majority had higher than expected (disproportionate increase) copy number in *M. glabribracteatum*. Other lower copy Ty1-*copia* repeats were found in similar proportions across all diploid species. Additionally, amounts of all Ty1-*copia* type repeats were in amounts proportional to the difference in genome sizes in *M. americanum* and *M. linearilobum*. Copy number variation in Ty3-*gypsy* lineages were also found but was less pronounced than the Ty1-*copia* retroelements (Fig. 4).

Several of the clusters were identified as satellite DNA repeats (Supplementary Table S7 available on Dryad). Most were found in at least two of the three analyzed diploid species, including the 4 nt microsatellite (referred to as satDNA1 in Supplementary Table S7 available on Dryad), and two repeats with monomer lengths of 155 and 180 nt (referred to as satDNA2 and satDNA3 in Supplementary Table S7 available on Dryad), respectively. The microsatellite and the 180 nt repeat were not found in *M. glabribracteatum*, while the 155 nt repeat was most prevalent in this species (Supplementary Table S7 available on Dryad) with much lower amounts in the other two diploids. All of these tandem repeats were also present in allopolyploids, albeit not necessarily in additive amounts (Supplementary Table S7 available on Dryad). Other satellite DNA repeats were represented in lower copy numbers.

The setup of the comparative analysis enabled a direct comparison of the genomic content of allopolyploids relative to their lower-ploid parental species. The allopolyploid genomes exhibited strong adherence to the additive expectation across all repeat types, albeit with a slight bias towards underrepresentation of some lineages (particularly of the Maximus/SIRE type; Fig. 5). The largest deviation from patterns expected under additivity was found in the allotetraploid *M. strigosum*, while the allohexaploids, *M. pringlei* and *M. sericeum*, had higher similarity to the immediate diploid and allotetraploid parents (Fig. 5).

DISCUSSION

Dating the Species Network

This article presents an indirect method for divergence time estimation on hybridization networks using the program BEAST2. This method is conceptually similar to the AlloppMUL model of Jones et al. (2013), where subgenomes of allopolyploids are treated as if they belonged to different species and can thus be analyzed within the multispecies coalescent framework

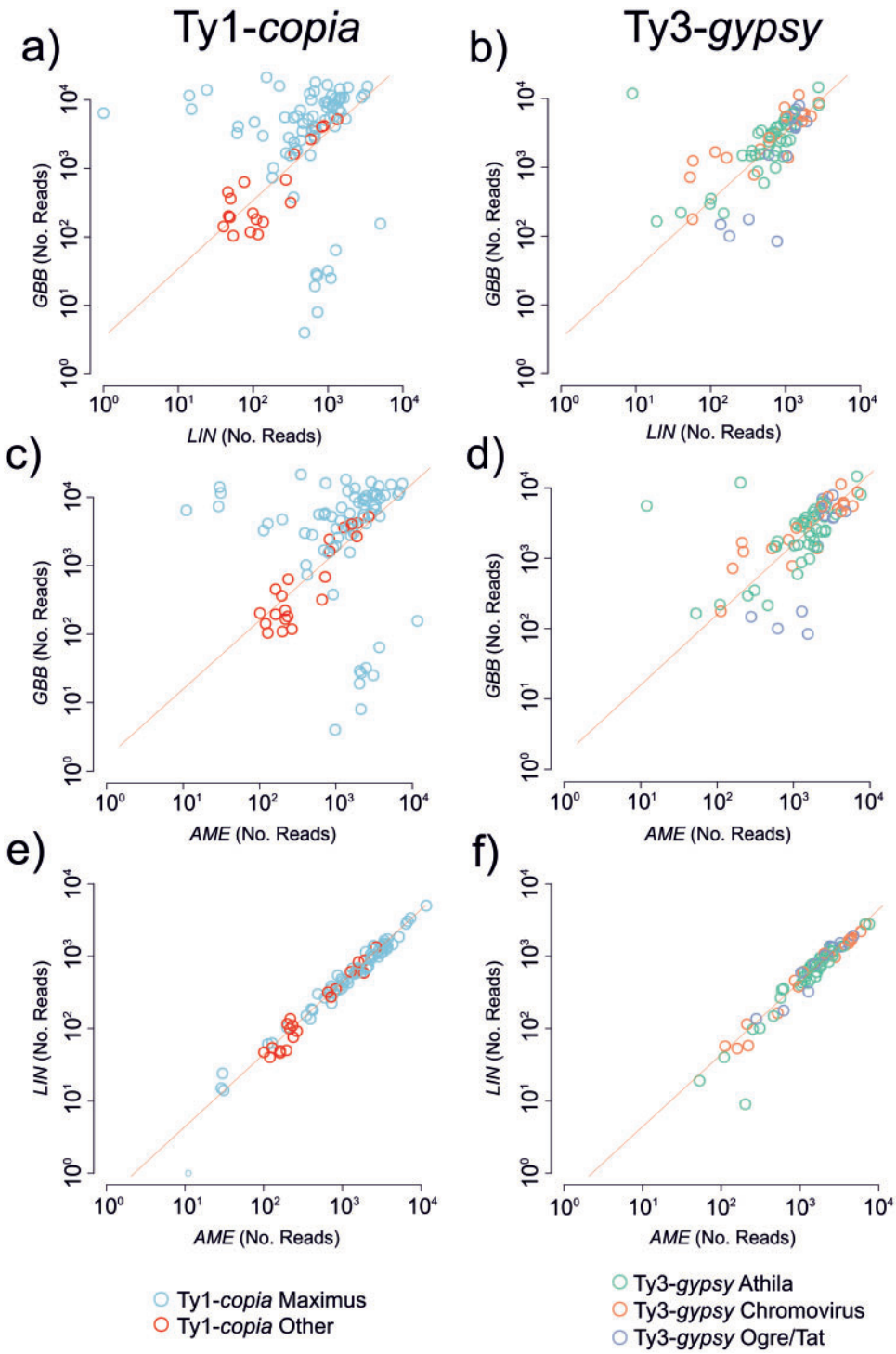

FIGURE 4. Pairwise scatterplots of the number of reads from each diploid species in repeat clusters from the comparative analysis. The slope of the red line is equal to the ratio of the genome sizes of the two species, thus, repeats (points) on the line are found in the same genomic proportions in species compared. a), c), e): Ty1-*copia* elements; b), d), f): Ty3-*gypsy* elements. The species are abbreviated as follows: AME = *M. americanum*; GBB = *M. glabribracteatum*; LIN = *M. linearilobum*.

(Heled and Drummond 2010). Using species trees for estimating allopolyploidization time has the advantage of taking lineage sorting and incomplete sampling of genes readily into account. This alleviates problems arising from differences in gene coalescence times (Doyle and Egan 2010; Kellogg 2016) and allows time calibrations to be placed on the internal nodes of species trees rather than the gene trees themselves. A technical advantage is that it can make use of available dating tools (such as BEAST) without having to resort to *ad*

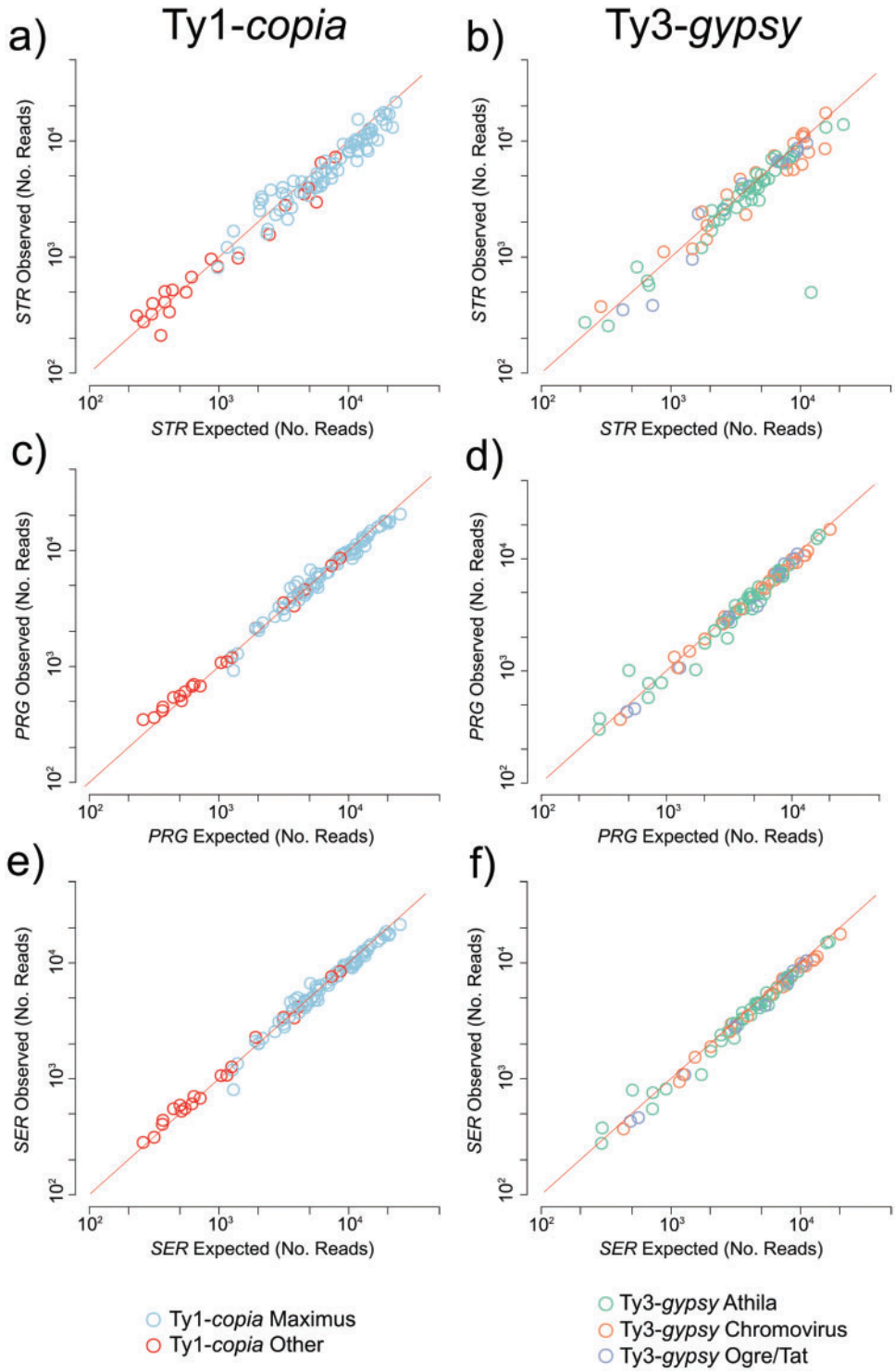

FIGURE 5. Pairwise scatterplots of the number of reads from each polyploid species and the sum of the parental taxa in repeat clusters from the comparative analysis. The slope of the red line is one, reflecting the expectation of additivity in allopolyploid genomes (i.e., the number of reads in allopolyploids should equal the sum of the number of reads in both parents). a, c, e): Ty1-*copia* elements; b, d, f): Ty3-*gypsy* elements. The species are abbreviated as follows: STR = *M. strigosum*; PRG = *M. pringlei*; SER = *M. sericeum*.

hoc approaches, such as the multistep approach used by Marcussen et al. (2015).

The cross-bracing method applied to species trees allows for variation in the coalescence times of genes in the ancestral populations of the children of the cross-braced nodes, while keeping the nodes in the species trees (nearly) the same. Also, the increased information on node split times (i.e., two nodes diverged at the same

time in different parts of the tree) may lead to better divergence time estimates across the whole tree (Shih and Matzke 2013). Additionally, alternative scenarios of hybridization can be directly addressed using model testing approaches via Bayes Factors (Baele et al. 2012, 2013). This applies to successive allopolyploidization events, as in our case, or to alternative scenarios for parentage of allopolyploids. However, this can become unwieldy, as the number of analyses that need to be run will quickly increase with uncertainty in allopolyploid origin.

Naturally, the cross-bracing method has some potential caveats. Homoeologous sequences have to be assigned to their respective parental genomes prior to the dating analysis, which will be problematic if measures, such as distances among sequences (as done here), give ambiguous results. These may be alleviated by integrating over assignment uncertainty as part of the analysis, as is done for the AlloppMUL and AlloppNET models (Jones et al. 2013, Jones 2017). The current implementation in these models, however, assumes a pair of homoeologues per tetraploid individual (no higher ploidy level implemented yet: Jones et al. 2013, Jones 2017), i.e., it is not explicitly taking allelic variation into account [although these can be resolved in an *ad hoc* manner, as explained in the manual to the AlloppNET model available from http://indriid.com/workingnotes2013.html (accessed 25 October 2017)].

Here, we assume that the allopolyploid parentage is known, or can be tested or reduced to a reasonably small number of candidates such that topology testing can be applied. In case one or both of the parental species are extinct or have remained unsampled, using a species closely related to the true parental species will likely lead to overestimation of the age of the allopolyploid (Doyle and Egan 2010). These issues are avoided by the AlloppNET model (Jones et al. 2013; Jones 2017), which estimates the ages of so-called hybridization nodes (i.e., the age of the allopolyploidization event), which are distinct from the nodes pertaining to the split between the subgenome of an allopolyploid and its parental species (Fig. 2 in Jones 2017). This method, however, cannot be applied at higher allopolyploid levels (Jones et al. 2013; Jones 2017).

Like other methods, including the AlloppNET model (Jones et al. 2013; Jones 2017), the cross-bracing method does not explicitly account for multiple origins, which is very common in natural plant populations (Soltis et al. 2010). This is not a problem *per se*, as ancestral allelic variation in the allopolyploid subgenomes due to multiple origin is readily accommodated by the multispecies coalescent. If multiple origins over an extended period of time are suspected or suggested by other evidence, these may be mimicked by allowing a larger difference between the cross-braced node ages.

A technical disadvantage of cross-bracing is the increased length of the MCMC chain that needs to be run to obtain good effective sample sizes (ESS) for the cross-braced nodes (Shih and Matzke 2013). The narrow prior calibrations on the difference between cross-braced nodes make it difficult for the operators on node heights to make successful proposals during the MCMC, i.e., any node change proposal on one of a group of cross-braced nodes is very likely to be rejected because of the narrow prior on the difference between the cross-braced nodes. The dates of cross-braced nodes can change, but they will be sampled more slowly, because, in effect, one node date can be changed, and is unlikely to move again until other cross-braced nodes "catch up." Therefore, only node dates which are not cross-braced (including the root) will sample quickly. This effect can be seen in BEAST runs with and without cross-bracing. In our analysis, about half of the nodes were cross-braced, and the conditional acceptance rate [Pr(acc | m) in the screen log of a BEAST run] was reduced from 0.1834 to 0.0717. This demonstrates the reduced efficiency of parameter-space exploration during the MCMC. This can be ameliorated by increasing the number of generations of the MCMC by a factor proportional to the number of cross-braced nodes. It is expected that designing a new operator in BEAST that moves the date of cross-braced nodes at once would result in dramatic improvements in proposal acceptance, and higher ESS/hour.

Allopolyploid Species Phylogeny of Melampodium sect. Melampodium

The age of the species phylogeny was calibrated using the divergence time estimates from the Heliantheae alliance. The inferred age of the whole genus *Melampodium* was determined to lie within the early to middle Miocene, while section *Melampodium* was placed in the late Pliocene to early Miocene (3.4–6.8 Ma).

The age estimates for the allopolyploids, ranging from 0.23 to 1.41 Ma (Fig. 2), suggest that they all formed during the Pleistocene. Allopolyploid formation with respect to age, climatic change, and harsher climates has been discussed in the literature (Brochmann et al. 2004). It has been proposed that allopolyploids may have a selective advantage in more variable habitats, perhaps implying that they may have been more likely to form and persist during periods of change. Mexico has not been exempt from such climatic fluctuations, as has been reviewed by Metcalfe et al. (2000).

In such periods of climatic fluctuations, successive lineage divergence may have been triggered in short intervals, causing difficulties in inferring the precise temporal order of species origin. This is evident for the allohexaploids *M. sericeum* and *M. pringlei*, where topology testing of the different scenarios potentially leading to the formation of the two allohexaploids (shared origin, *M. pringlei* first, and *M. sericeum* first) provided no decisive support for any of the scenarios (Bayes Factors <1). Support for an independent origin of the two species from recurrent hybridization of parental taxa in line with commonly observed multiple origins of allopolyploids (Soltis et al. 2010) is provided from divergent trajectories of rDNA sequence and loci

evolution (Weiss-Schneeweiss et al. 2012). In contrast to rDNA evolution, however, the overall repetitive DNA composition is remarkably similar in the two allohexaploids. Evidently, more data, also at the population level, will be necessary to elucidate the details of allopolyploid history of these species.

Repetitive DNA Evolution in Melampodium sect. Melampodium

Despite having the same number of chromosomes, the three diploid parental taxa of the *Melampodium* allopolyploids have disparate genome sizes ranging from 0.49 (*M. linearilobum*) to 1.85 pg/1C (*M. glabribracteatum*). As the genome size of *M. americanum* is similar to the inferred ancestral genome size of the whole section *Melampodium* (McCann 2017), larger and smaller genome sizes in *M. glabribracteatum* and *M. linearilobum*, respectively, represent both major trends proposed: the up- and downsizing of plant genomes during evolution (Lysak et al. 2009). Differential accumulation and/or deletion of a small number of high-abundance repeat families, as suggested for a number of related diploid species (Ty3-*gypsy* Ogre in *Vicia*, Macas et al. 2015; Ty3-*gypsy* Gorge 3 in *Gossypium*, Hawkins et al. 2009), is found in *M. glabribracteatum*, where preferential amplification of the Maximus lineage of the Ty1-*copia* retrotransposons led to an increase of genome size (Table 2). In contrast, proportional changes in copy numbers across the majority of repeats, as suggested for the giant genomes of *Fritillaria* (Kelly et al. 2015), are in line with the similar relative proportions of all major repeat types in *M. americanum* and *M. linearilobum* (Fig. 4). *Melampodium linearilobum* experienced genome downsizing since its divergence from other species in this series, 0.5 to 1.5 Ma.

The genome sizes of *M. strigosum* (4*x*), *M. pringlei* (6*x*), and *M. sericeum* (6*x*) were additive (or nearly so) in comparison to the extant relatives of their parental taxa (Weiss-Schneeweiss et al. 2012), which is suggestive of little to no change in genome size following polyploidization. Such genome stasis is well-supported in this study by the roughly commensurate genomic proportions of most repeat types in allohexaploids and that expected from the parental genomes (Fig. 5). The lack of significant restructuring of the repeatome is also reflected in the efficiency of GISH. The parental and grandparental (in the allohexaploids) subgenomes in the allopolyploids were unequivocally labeled (Fig. 3, Supplementary Fig. S8 in Appendix S4 available on Dryad) indicating low levels of cross-subgenome repetitive DNA homogenization (Lim et al. 2007; Renny-Byfield et al. 2013; Dodsworth et al. 2017).

General repetitive DNA and genome size additivity, however, does not have to imply a complete lack of genome turnover. Factors driving change following polyploidization and speciation events may have acted on lower copy number sequences in these allopolyploids, such as the identified tandem repeats. Differential evolution of rDNA loci, including cytological diploidization and sequence evolution have already been demonstrated in both the allotetraploid and its allohexaploid derivatives (Weiss-Schneeweiss et al. 2012). Accordingly, this study shows that satellite DNA and rDNA repeats do show some deviation from additivity, which due to their generally lower copy number may not be reflected in total genome size changes. However, it is known that such repeats exhibit relatively fast rates of turnover including change in chromosomal localization, copy number and monomer type (Garrido-Ramos 2015). Therefore, a more fine-grained analysis of these tandem repeats, including localization of satellite DNAs in the chromosomes using fluorescent *in situ* hybridization (FISH), will be necessary to confidently test the presence and extent of genomic stasis in the *Melampodium* allopolyploids.

The relatively recent origin of the allopolyploids analyzed in this study (<1.4 Ma) likely played a role in the absence of more significant changes in the repetitive fraction of their genomes. Although changes in this allopolyploid complex were relatively low compared to those found in other systems (Renny-Byfield et al. 2012; Dodsworth et al. 2017), differences in the levels of deviation from additivity with respect to ploidy level were detectable. Relative to its allohexaploid descendants, the allotetraploid *M. strigosum* displayed an overall trend of increased disparity from additivity (biased towards underrepresentation) across the majority of repeats. This was particularly the case in repeats of Ty1-*copia* Maximus type, although, due to low solo-LTR to full element ratios (Table 3), cannot be explained by intrastrand recombination alone. In the established temporal framework used here, the observed patterns can be explained by the necessity that *M. strigosum* is older and thus, in the absence of the more typically observed punctuated transposable element evolution following polyploid formation (Parisod and Senerchia 2012; Bennetzen and Wang 2014), is likely to exhibit more change overall than its derivative species.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.dg8q0.

Funding

This work was supported by Austrian Science Fund (FWF) [project P25131 to H.W.S.]; Czech Science Foundation [BP501/12/G090 to J.M.] and Czech Academy of Sciences [RVO:60077344 to J.M.]; Discovery Early Researcher Award DE150101773, awarded by the Australian Research Council to N.J.M.

Acknowledgments

The authors acknowledge financial support of the Austrian Science Fund (FWF), Czech Science

Foundation, and Czech Academy of Sciences. Access to computing and storage facilities owned by the Vienna Scientific Cluster (Vienna, Austria) and Czech National Grid Infrastructure MetaCentrum provided under the program "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated. We thank Enrique Ortiz (UNAM, Mexico) for logistic and physical help during the fieldtrip to collect plant material. We would also like to thank the editors and anonymous reviewers for insightful comments on previous versions of this manuscript.

References

Ayres D.L., Darling A., Zwickl D.J., Beerli P., Holder M.T., Lewis P.O., Huelsenbeck J.P., Ronquist F., Swofford D.L., Cummings M.P., Rambaut A., Suchard M.A. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst. Biol. 61:170–173.

Baele G., Lemey P., Bedford T., Rambaut A., Suchard M.A., Alekseyenko A.V. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29:2157–2167.

Baele G., Li W.L.S., Drummond A.J., Suchard M.A., Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. Mol. Biol. Evol. 30:239–243.

Barker M.S., Baute G.J., Liu S.L., 2012. Duplications and turnover in plant genomes. In: Wendel J., Greilhuber J., Dolezel J., Leitch I.J. Plant genome diversity, vol. 1. Vienna, Austria: Springer, p. 155–169.

Bennetzen J.L., Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu. Rev. Plant Biol. 65:505–530.

Bertrand Y.J., Scheen A.-C., Marcussen T., Pfeil B.E., de Sousa F., Oxelman B. 2015. Assignment of homoeologs to parental genomes in allopolyploids for species tree inference, with an example from *Fumaria* (Papaveraceae). Syst. Biol. 64:448–471.

Blöch C., Weiss-Schneeweiss H., Schneeweiss G.M., Barfuss M.H., Rebernig C.A., Villaseñor J.L., Stuessy T.F. 2009. Molecular phylogenetic analyses of nuclear and plastid DNA sequences support dysploid and polyploid chromosome number changes and reticulate evolution in the diversification of *Melampodium* (Millerieae, Asteraceae). Mol. Phylogenet. Evol. 53:220–233.

Brochmann C., Brysting A.K., Alsos I.G., Borgen L., Grundt H.H., Scheen A.C. Elven R. 2004. Polyploidy in arctic plants. Biol. J. Linnean Soc. 82:521–536.

Chester M., Gallagher J.P., Symonds V.V., da Silva A.V.C., Mavrodiev E.V., Leitch A.R., Soltis P.S., Soltis D.E. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proc. Natl. Acad. Sci. USA 109:1176–1181.

Chester M., Riley R., Soltis P., Soltis D. 2015. Patterns of chromosomal variation in natural populations of the neoallotetraploid *Tragopogon mirus* (Asteraceae). Heredity 114:309–317.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6:836–846.

Dodsworth S., Leitch A.R., Leitch I.J. 2015. Genome size diversity in angiosperms and its influence on gene space. Curr. Opin. Genet. Dev. 35:73–78.

Dodsworth S., Jang T.S., Struebig M., Chase M.W., Weiss-Schneeweiss H., Leitch A.R. 2017. Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Plant Syst. Evol. 303:1013–1020.

Doyle J., Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.

Doyle J.J., Egan A.N. 2010. Dating the origins of polyploidy events. New Phytol. 186:73–85.

Drummond A.J., Ho S.Y., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:699.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.

Garrido-Ramos M.A. 2015. Satellite DNA in plants: more than just rubbish. Cytogenet. Genome Res. 146:153–170.

Grant V. 1981. Plant speciation. New York, USA: Columbia University Press.

Hawkins J.S., Proulx S.R., Rapp R.A., Wendel, J.F. 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc. Natl. Acad. Sci. USA 106:17811–17816.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Hollister J.D. 2015. Polyploidy: adaptation to the genomic environment. New Phytol. 205:1034–1039.

Huang C.-H., Zhang C., Liu M., Hu Y., Gao T., Qi J., Ma H. 2016. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. Mol. Biol. Evol. 33:2820–2835.

Jang T.-S., Weiss-Schneeweiss H. 2015. Formamide-free genomic *in situ* hybridization allows unambiguous discrimination of highly similar parental genomes in diploid hybrids and allopolyploids. Cytogenet. Genome Res. 146:325–331.

Jiao Y.N., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho P.E., Hu Y., Liang H.Y., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H., Leebens-Mack J., dePamphilis C.W. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100.

Jones G. 2017. Bayesian phylogenetic analysis for diploid and allotetraploid species networks. bioRxiv, doi: 10.1101/129361.

Jones G., Sagitov S., Oxelman B. 2013. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. Syst. Biol. 62:467–478.

Kay K.M., Whittall J.B., Hodges S.A. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol. Biol. 6:36.

Kellogg E.A. 2016. Has the connection between polyploidy and diversification actually been tested? Curr. Opin. Plant Biol. 30:25–32.

Kelly L.J., Renny-Byfield S., Pellicer J., Macas J., Novák P., Neumann P., Lysak M.A., Day P.D., Berger M., Fay M.F., Nichols R.A., Leitch A.R., Leitch I.J. 2015. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol. 208:596–607.

Kim K.-J., Choi K.-S., Jansen R.K. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Mol. Biol. Evol. 22:1783–1792.

Koh J., Soltis P.S., Soltis D.E. 2010. Homoeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). BMC Genomics 11:97.

Kovařík A., Dadejova M., Lim Y.K., Chase M.W., Clarkson J.J., Knapp S., Leitch A.R. 2008. Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. Ann. Bot. 101:815–823.

Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33:1870–1874.

Leitch I., Bennett M. 2004. Genome downsizing in polyploid plants. Biol. J. Linnean Soc. 82:651–663.

Lim K.Y., Kovařík A., Matyasek R., Chase M.W., Clarkson J.J., Grandbastien M., Leitch A.R. 2007. Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytol. 175:756–763.

Lysak M.A., Koch M.A., Beaulieu J.M., Meister A., Leitch I.J. 2009. The dynamic ups and downs of genome size evolution in Brassicaceae. Mol. Biol. Evol. 26:85-98.

Ma X.-F., Gustafson J. 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. Cytogenet. Genome Res. 109:236–249.

Macas J., Meszaros T., Nouzova M. 2002. PlantSat: a specialized database for plant satellite repeats. Bioinformatics 18:28–35.

Macas J., Novák P., Pellicer J., Čížková J., Koblížková A., Neumann P., Fuková I., Doležel J., Kelly L.J., Leitch I.J. 2015. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe *Fabeae*. PLoS One 10:e0143424.

Madlung A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110:99–104.

Mandáková T., Kovařík A., Zozomová-Lihová J., Shimizu-Inatsugi R., Shimizu K.K., Mummenhoff K., Marhold K., Lysak M.A. 2013. The more the merrier: recent hybridization and polyploidy in *Cardamine*. Plant Cell 25:3280–3295.

Mandáková T., Marhold K., Lysak M.A. 2014. The widespread crucifer species *Cardamine flexuosa* is an allotetraploid with a conserved subgenomic structure. New Phytol. 201:982–992.

Marcussen T., Jakobsen K.S., Danihelka J., Ballard H.E., Blaxland K., Brysting A.K., Oxelman B. 2012. Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, Violaceae). Syst. Biol. 61:107–126.

Marcussen T., Heier L., Brysting A.K., Oxelman B., Jakobsen K.S. 2015. From gene trees to a dated allopolyploid network: insights from the angiosperm genus *Viola* (Violaceae). Syst. Biol. 64:84–101.

Mayrose I., Zhan S.H., Rothfels C.J., Magnuson-Ford K., Barker M.S., Rieseberg L.H., Otto S.P. 2011. Recently formed polyploid plants diversify at lower rates. Science 333:1257–1257.

McCann J., Schneeweiss G.M., Stuessy T.F., Villaseñor J.L., Weiss-Schneeweiss H. 2016. The impact of reconstruction methods, phylogenetic uncertainty and branch lengths on inference of chromosome number evolution in American daisies (*Melampodium*, Asteraceae). PLoS One 11:e0162299.

McCann J. 2017. Genome evolution of diploids and polyploids in genus *Melampodium* (Asteraceae) [Ph.D. Thesis]. University of Vienna. p. 39–66.

Metcalfe S.E., O'Hara S.L., Caballero M., Davies S.J. 2000. Records of Late Pleistocene–Holocene climatic change in Mexico—a review. Quat. Sci. Rev. 19:699–721.

Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-Tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Novák P., Neumann P., Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11:378.

Novák P., Neumann P., Pech J., Steinhaisl J., Macas J. 2013. RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793.

Ogilvie H.A., Drummond A.J. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Mol. Biol. Evol. 34:2101–2114.

Parisod C., Senerchia N. 2012. Responses of transposable elements to polyploidy. In: Grandbastien M.A., Casacuberta J.M., editors. Plant transposable elements. Vienna, Austria: Springer. pp. 147–168.

Renny-Byfield S., Chester M., Kovařík A., Le Comber S.C., Grandbastien M.-A., Deloger M., Nichols R., Macas J., Novák P., Chase M.W., Leitch A.W. 2011. Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. Mol. Biol. Evol. 28:2843–2854.

Renny-Byfield S., Chester M., Nichols R.A., Macas J., Novák P., Leitch A.R. 2012. Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. PLoS One 7:e36963.

Renny-Byfield S., Kovařík A., Kelly L.J., Macas J., Novák P., Chase M.W., Nichols R.A., Pancholi M.R., Grandbastien M.-A., Leitch A.R. 2013. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low-and high-copy sequences. Plant J. 74:829–839.

Rieseberg L.H., Willis J.H. 2007. Plant speciation. Science 317:910–914.

Shih P.M., Matzke N.J. 2013. Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. Proc. Natl. Acad. Sci. USA 110:12355–12360.

Soltis D.E., Buggs R.J., Doyle J.J., Soltis P.S. 2010. What we still don't know about polyploidy. Taxon 59:1387–1403.

Stuessy T.F., Blöch C., Villaseñor J.L., Rebernig C.A., Weiss-Schneeweiss H. 2011. Phylogenetic analyses of DNA sequences with chromosomal and morphological data confirm and refine sectional and series classification within *Melampodium* (Asteraceae, Millerieae). Taxon 60:436–449.

Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9:1.

Torices R. 2010. Adding time-calibrated branch lengths to the Asteraceae supertree. J. Syst. Evol. 48:271–278.

Weiss-Schneeweiss H., Blöch C., Turner B., Villaseñor J.L., Stuessy T.F., Schneeweiss G.M. 2012. The promiscuous and the chaste: frequent allopolyploid speciation and its genomic consequences in American daisies (*Melampodium* sect. *Melampodium*; Asteraceae). Evolution 66:211–228.

Weiss-Schneeweiss H., Emadzade K., Jang T.-S., Schneeweiss G. 2013. Evolutionary consequences, constraints and potential of polyploidy in plants. Cytogenet. Genome Res. 140:137–150.

Wendel J.F. 2015. The wondrous cycles of polyploidy in plants. Am. J. Bot. 102:1753–1756.

Wolfe K.H. 2001. Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. 2:333–341.

Wood T.E., Takebayashi N., Barker M.S., Mayrose I., Greenspoon P.B., Rieseberg L.H. 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA 106:13875-13879.

Zozomová-Lihová J., Mandáková T., Kovaříková A., Mühlhausen A., Mummenhoff K., Lysak M.A., Kovařík A. 2014. When fathers are instant losers: homogenization of rDNA loci in recently formed *Cardamine schulzii* trigenomic allopolyploid. New Phytol. 203:1096–1108.